Abstract — PieceWise Affine (PWA) model comprises several affine dynamics defined over polyhedral regions in the regressor (state+input) space. Identification of a PWA model is very often a starting point for the controller synthesis of hybrid systems. In this paper we extend the clustering-based procedure for identification of a PieceWise AutoRegressive eXogenous (PWARX) model proposed in [Ferrari-Trecate et al., 2003]. By exploiting a priori process knowledge we choose an appropriate linear transformation of the regression vector for a better and more efficient identification of the process nonlinearities. We significantly reduce the computational complexity of the classification algorithm for finding the complete polyhedral partition of the model domain. This modified clustering-based procedure is used to identify a PWARX model of the electronic throttle - a highly nonlinear component that regulates air inflow to the engine of a car.

I. INTRODUCTION

Electronic throttle is a DC motor driven valve that regulates air inflow into the engine’s combustion system. It is burdened with strong nonlinear effects of friction and return spring characteristics, both of which make the controller synthesis a challenging task. Various strategies for identification and control of the electronic throttle have been proposed in the literature in the past, cf. [1], [2], [3].

We focus on the optimal control strategy based on the PWA throttle model described in [4]. PWA models are a natural extension of linear models that can describe and/or approximate nonlinear processes. Furthermore, as recently shown in [5], for this class of models one can explicitly solve constrained optimal control problems (i.e. compute the control law off-line). Unfortunately, the PWA model used in [4] is hard to obtain since it demands detailed knowledge on the underlying physical behavior of the throttle.

One of few algorithms for PWA identification that allow discontinuous PWA functions is the clustering-based procedure for the identification of PieceWise AutoRegressive eXogenous (PWARX) models [6] that combines clustering, linear identification and pattern recognition techniques. It starts from the assumption that there exists an optimal PWARX model for the process in question and that regression vectors that are close to each other are likely to belong to the same ARX submodel of the optimal PWARX model. The identification procedure described in this paper should be seen as an extension of the procedure in [6] in a way that makes it more efficient and tractable for a broader class of systems.

Local ARX submodels are first estimated around each data point (local regression), together with the corresponding measures of confidence. Then similar local submodels estimated around data points close in the regressor space are clustered using the K-means algorithm [7] enhanced with the measures of confidence to improve the clustering performance. After the original data has been classified, the final affine submodels and the polyhedral regions on which they are valid are estimated, using weighted least squares and linear classification, respectively. In order to successfully apply the procedure to an electronic throttle, two significant modifications have to be made. The first one is a linear transformation of the regression vector prior to the local regression, using a priori process knowledge, where past inputs and outputs are approximately transformed into the states of the process along which process nonlinearities are mostly emphasized. The second one is a computational complexity reduction of Multicategory Robust Linear Programming (MRLP) [8], one of the linear classification algorithms, that is used to find the hyperplanes separating identified clusters of regression vectors, using only the vertices of each cluster, instead of all the points. The separating hyperplanes form convex polyhedral regions in the regressor space on which different affine submodels are valid.

The paper is organized as follows. In Section II the PWARX model is defined and in Section III the existing clustering-based procedure for the PWARX model identification is summarized. In Section IV the electronic throttle is briefly described. The modifications of the identification procedure are discussed in Sections V and VI, and the resulting discrete-time PWA electronic throttle model is presented and validated in Section VII.

II. PWARX MODEL

A PWA map $f(\cdot) : \mathbb{X} \rightarrow \mathbb{R}$ is defined as:

$$f(x) = \begin{cases} \theta_1^T \begin{bmatrix} x \\ 1 \end{bmatrix} , & \text{if } x \in \mathcal{X}_1, \\ \vdots \\ \theta_s^T \begin{bmatrix} x \\ 1 \end{bmatrix} , & \text{if } x \in \mathcal{X}_s, \end{cases}$$

(1)
where $X \subset \mathbb{R}^n$ is a bounded polyhedron referred to as the regressor space, $x$ is a regression vector defined later, $\theta_i \in \mathbb{R}^{n+1}$, $i = 1, \ldots, s$, are parameter vectors (PVs), $s$ is the number of submodels and $X_i$, $i = 1, \ldots, s$, are convex polyhedral regions in the regressor space $X$. The polyhedral regions $X_i$ form the partition $\{X\}_{i=1}^s$ of the regressor space $X$, i.e.

$$\bigcup_{i=1}^s X_i = X, \quad X_i \cap X_j = \emptyset, \quad i \neq j \quad (2)$$

PWARX models are of the form:

$$y_k = f(x_k) + e_k, \quad (3)$$

where $k$ is the time index, $e_k$ is the model prediction error term, and $x_k \in \mathbb{R}^n$ is a regression vector defined as:

$$x_k = [y_{k-1} \ y_{k-2} \ \cdots \ y_{n_k} u_{k-1} \ u_{k-2} \ \cdots \ u_{n_k}]^T, \quad (4)$$

where $y_k \in \mathbb{R}$ is the output and $u_k \in \mathbb{R}^m$ the input of the process. Note that $n = n_a + mn_b$, where $n_a$ and $n_b$ denote the number of past inputs and outputs used in the regression vector. As it will be shown later, instead of (4) we use a linearly transformed regression vector to define a PWARX model (3).

III. Clustering-based identification procedure

In this section the clustering-based identification procedure is shortly described. A more detailed description can be found in [6]. The PWA identification problem can be stated as a reconstruction of the PWA map $f(\cdot)$ from a given dataset $\mathcal{N} = \{(x_k, y_k)\}_{k=1}^N$. As in the most of PWA identification procedures, model orders $n_a$ and $n_b$ as well as the number of submodels $s$ are fixed in advance. To make a good initial choice of these parameters a priori process knowledge is useful. Their most appropriate values can be chosen in an iterative procedure, based on validation of the resulting model, thus enabling one to weight between model fitting and model complexity. The clustering-based identification procedure consists of the following four main steps:

1. Local Regression. Local Datasets (LDs) $\mathcal{C}_k$, $k = 1, \ldots, N$ are created by collecting a point $(x_k, y_k)$ and its $c-1$ nearest neighbors (according to the Euclidian distance in the regressor space), i.e. the data points $(\bar{x}, \bar{y})$ that satisfy

$$||x_k - \bar{x}||^2 \leq ||x_k - \tilde{x}||^2, \quad \forall(\bar{x}, \bar{y}) \in \mathcal{N} \setminus \mathcal{C}_k \quad (5)$$

where parameter $c$ is a tuning “knob” of the clustering-based procedure. Assuming that there exists an exact PWA mapping $f$ that describes the process best, LDs that contain data belonging to similar submodels of $f$ are referred to as pure LDs, while those containing data generated by substantially different affine submodels are called mixed LDs. This distinction is conceptual because the true regions are unknown, especially at this stage of the algorithm.

For each LD a local PV $\theta^{L,S,k}$ is estimated using least squares on $\mathcal{C}_k$. Local parameter vector $\theta^{L,S,k}$, together with the mean $m_k = \frac{1}{s} \sum_{x \in \mathcal{C}_k} x$ of the regression vectors in $\mathcal{C}_k$, forms the feature vector $\xi_k = [(\theta^{L,S,k})^T, m_k^T]^T$. Feature vectors $\xi_k$ allow the reconstruction of different submodels that share similar coefficients but are defined on distant regions of the regressor space. Distant regions should not be unified in a single convex polytope, since they are likely to cross through regions of other submodels. Additional confidence measure $R_k^{-1}$ of a feature vector $\xi_k$ is calculated based on the empirical covariance matrices of the PVs and the scatter matrices of the regression vectors in each LD.

2. Clustering. Feature vectors $\xi_k$, $k = 1, \ldots, N$ are clustered in $s$ clusters $\mathcal{D}_i$ by minimizing a suitable cost function. This is performed using a modified K-means algorithm [7] which exploits the confidence measures $R_k^{-1}$ to avoid local minima of the cost function as well as to reduce sensitivity to outliers.

3. Estimation of submodels. By using the bijective maps

$$(x_k, y_k) \longleftrightarrow \mathcal{C}_k \longleftrightarrow \xi_k \quad (6)$$

data subsets $\mathcal{F}_i$, $i = 1, \ldots, s$ are built according to clusters $\mathcal{D}_i$. The corresponding PVs $\theta_i$ are estimated using weighted least squares on data subsets $\mathcal{F}_i$. The confidence measures used are derived from matrices $R_k^{-1}$.

4. Estimation of the regions. The convex polyhedral regions $X_i$, $i = 1, \ldots, s$, can be found in two different ways. The first one is to find a hyperplane that separates the convex hull of the regression vectors $x$ belonging to $\mathcal{F}_i$ from that of the $\mathcal{F}_j$, for any pair $(i, j)$, $i \neq j$. This is accomplished by employing a linear classifier like Support Vector Machine (SVM) [9] with a linear kernel, to solve a linear pattern recognition problems. If the subsets of the regression vectors are not linearly separable SVM minimizes the error due to misclassified vectors. However, a major drawback of this method is that, when the regressor space $X$ has more than one dimension, and the number of submodels (classes) is larger than two ($s > 2$), the estimated regions might not form a partition of $X$, i.e. the “holes” in the regressor space may exist, where regression vectors do not belong to any region.

The second way of finding polyhedral regions $X_i$ is by solving a multiclassification classification problem involving all the available data at once. One such method is MRLP [8] and it always results in a partition of $X$ satisfying (2), i.e. without “holes”, using a single linear program. However, computational complexity is high, mostly because of the memory requirements for storing the linear program matrices, and is thus applicable only when the number of regression vectors (amount of experimental data) and the number of the submodels are relatively small. This problem is further discussed in section VI where we show how MRLP, simplified under reasonable assumptions, results in suboptimal but computable separating hyperplanes that still define the partition of $X$ without “holes”.

IV. Electronic throttle

Electronic throttle is a DC motor driven valve that regulates air inflow into the engine’s combustion system. The
DC motor is controlled using the control voltage $u \,[V]$, and the air inflow is maintained by certain valve plate position (angle) $y = \Theta \,[^{\circ}]$, with the return spring attached to the plate’s shaft that counterbalances the motor torque. This device replaces its classical mechanical counterpart and brings several benefits regarding driving performance, fuel consumption and vehicle emission. However, it is burdened with two strong nonlinear effects: friction with presliding effect [3] and return spring characteristics, both of which make the controller synthesis a challenging task. The nonlinearity of the stress-strain characteristics of the return spring is introduced to place the throttle valve in a so-called Limp-Home (LH) position (denoted with $\Theta_{LH}$) when power supply fails. Such valve position enables the driver to “limp” to the nearest repairing facility since there is some air inflow into the car engine. The LH nonlinearity is shown in enlarged detail around the LH position in Fig. 1.

A discrete-time PWA model of the throttle can be derived by sampling linearized dynamics of the corresponding continuous-time nonlinear model, cf. [4]. However, significant discrepancy in their behavior may occur, since the sampling time of the discrete-time PWA model is required to be as large as possible and the switchings between its affine dynamics may occur only at multiples of the sampling time. To alleviate the problem additional affine dynamics should be introduced in the PWA model to predict switchings between the sampling instants as done in [4], but this often leads to a large number of submodels. On the other hand, using the altered clustering-based procedure we bypass this problem and derive a simple discrete-time PWA model that captures the process behavior well in a computationally efficient way, directly from experimental data.

V. Linear Transformation of the Regression Vector

Basic assumption of the clustering-based procedure is that regression vectors that are close to each other in the regressor space are likely to belong to the regions of $f$ with similar ARX models. Validity of this assumption depends on the true nature of the nonlinearities, the numerical values of the regressor components as well as the sampling schedule in the bounded regressor space $X$ and the choice of the parameter $c$. In other words, the outcome ratio between mixed and pure LDs after local regression should be small $^1$. The numerical values of the regressor components crucially influence the Euclidian distance between regression vectors, often causing improper grouping in LDs. Concerning the sampling schedule it is assumed that the input is persistently exciting and that the regression vectors are not all concentrated around the boundary of the sets $X_i$ of $f$, but for some nonlinearities this can only be achieved by gathering a very large set of data points if regression vector of the form (4) is used. Figure 2 illustrates the problem for the friction nonlinearity in the throttle example. In this case the line that marks zero angular velocity of the parameter $c$. In other words, the outcome ratio between mixed and pure LDs after local regression should be small $^1$. The numerical values of the regressor components crucially influence the Euclidian distance between regression vectors, often causing improper grouping in LDs. Concerning the sampling schedule it is assumed that the input is persistently exciting and that the regression vectors are not all concentrated around the boundary of the sets $X_i$ of $f$, but for some nonlinearities this can only be achieved by gathering a very large set of data points if regression vector of the form (4) is used. Figure 2 illustrates the problem for the friction nonlinearity in the throttle example. In this case the line that marks zero angular velocity

\[ \omega_{k-1} \sim \Theta_{k-1} - \Theta_{k-2} = 0 \] characterizes the boundary of different submodels due to friction. Many regression vectors that are on both sides of this line are closer (vectors inside the circle marked $C_1$) than just those on one side of the same line (vectors inside the circle marked $C_2$). The number of regression vectors would have to be very large in order to keep the ratio between mixed and pure LDs reasonably low, causing increased computational complexity. Hence, knowing that the nonlinearity of friction is emphasized along the state of angular velocity, introducing it into

\footnote{Assuming that only noise causes mixed LDs, the conditions of optimal assignment of data points to the submodel that most likely generated it have been examined in [10].}
The regression vector helps to group a larger number of pure LDs from a smaller set of collected data points. Since friction with emphasized presliding effect also depends on the plate acceleration \((\alpha_{k-1} \sim \Theta_{k-1} - 20\Theta_{k-2} + \Theta_{k-3})\) it is introduced in the regression vector as well.

In general, a priori process knowledge is combined with the identification procedure by performing a linear transformation of the regression vector, where past inputs and outputs are approximately transformed into states of the process along which the process nonlinearities are mostly emphasized. The linear transformation is defined as:

\[
\hat{x} = Lx,
\]

where \(L \in \mathbb{R}^{n \times n}\) and \(\hat{x} \in \hat{X}\) is the transformed regression vector. Assuming that the matrix \(L\) is of full rank, the final model can be transformed to the original PWARX form described in Section II by using the \(L^{-1}\) transformation.

Using one past input value in the regression vector, the linear transformation \(L\) used for the PWA identification of the electronic throttle is defined as:

\[
L = \begin{bmatrix}
1 & 0 & 0 & 0 \\
K_\nu & -K_\nu & 0 & 0 \\
K_\alpha & -2K_\alpha & K_\alpha & 0 \\
0 & 0 & 0 & K_u
\end{bmatrix},
\]

where the coefficients \(K_\nu, K_\alpha\) and \(K_u\) are scaling factors for dilating/contracting distances in the regressor space \(\hat{X}\). Dilation/contraction of distances is a form of linear transformation referred to as homothety, which is in this case used to adapt the distances in the regressor space resulting in a more efficient local regression. A general choice of the scaling factors is not covered in this paper and the practical advice would be to start with such factors that normalize the span of each component of the regression vector. The factors are then iteratively adjusted by trial and error, until a satisfying clustering is achieved, having in mind the a priori qualitative knowledge of the nonlinearity characteristic.

VI. REDUCING COMPUTATIONAL COMPLEXITY OF MRLP

Having assigned the data points to subsets \(F_i, i = 1, \ldots, s\), convex polyhedral regions \(X_i\) are built using separating hyperplanes between regression vectors in \(F_i\) from those in \(F_j, i \neq j\). A separating hyperplane in the regressor space \(X\) for any pair \((i, j), i < j\), is described by:

\[
M_{ij}x = m_{ij}, \quad \|M_{ij}\| \neq 0,
\]

where \(M_{ij} \in \mathbb{R}^n\) is a row vector and \(m_{ij} \in \mathbb{R}\). In a linearly separable case the following must be satisfied:

\[
x_k \in X_i \Rightarrow M_{ij}x_k \leq m_{ij}, \quad \text{(10)}
\]

\[
x_k \in X_j \Rightarrow M_{ij}x_k > m_{ij}.
\]

In a linearly inseparable case the error due to misclassified regression vectors is minimized.

The algorithm for the estimation of separating hyperplanes that can solve the classification problem with more than two data classes without “holes” as in (2), MRLP [8], is a linear program defined by:

\[
\begin{align*}
\min_{M_i, m_i, v_{ij}} & \sum_{i=1}^{s} \sum_{j=1}^{s} \frac{1}{N_{E_i}} \Sigma_{ij} \\
\text{s.t.} & \quad v_{ij} \geq -A_i(M_i - M_j)^T + (m_i - m_j)1_i + 1_i, \\
& \quad v_{ij} \geq 0, ~ i \neq j, ~ i, j = 1, \ldots, s,
\end{align*}
\]

where \(v_{ij}\) are slack variables, \(1_i\) is a vector of ones of the appropriate dimension, \(N_i\) is the cardinality of \(F_i\) and \(A_i\) is a matrix whose rows are the regression vectors in \(F_i\). The final coefficients of the separating hyperplanes are defined by \(M_{ij} = M_i - M_j\) and \(m_{ij} = m_i - m_j\). The equivalent linear program in the form

\[
\begin{align*}
\min (f^{LP})^T z \\
\text{s.t.} & \quad A^{LP}z \leq b^{LP}
\end{align*}
\]

is composed of a matrix \(A^{LP}\) with \(2Ns(s-1)\) rows and \(Ns(s+1) + (n+1)s\) columns. Memory requirements grow polynomially with the number of data points \(N\) and the number of submodels \(s\).

The criterion to be minimized in (13) is in the end affected only by the points that are misclassified, so the majority of points does not increase the criterion function at all. Assuming that the clustering is performed suitably, without outliers, the computational complexity of the classification can be fairly reduced by using only the vertices of each cluster of regression vectors, while the obtained resulting partition is similar to the optimal one. The proposed reduced linear program is defined by:

\[
\begin{align*}
\min_{M_i, m_i, v_{ij}} & \sum_{i=1}^{s} \sum_{j=1}^{s} \frac{1}{N_{E_i}} \Sigma_{ij} \\
\text{s.t.} & \quad v_{ij} \geq -E_i(M_i - M_j)^T + (m_i - m_j)1_i + 1_i, \\
& \quad v_{ij} \geq 0, ~ i \neq j, ~ i, j = 1, \ldots, s,
\end{align*}
\]

where \(E_i\) is a matrix whose rows are vertices of the convex hull formed by regression vectors in \(F_i\) and \(N_{E_i}\) is the number of rows of \(E_i\). The vertices of a convex hull can be found for example by a series of linear programs [11], computational complexity and memory requirements of which are negligible compared to the ones of the original MRLP. In Figure 3 comparison of MRLP classification results (separating lines) with all the points and with vertices only is depicted. The exemplifying classification problem is two-dimensional with 3 classes. It is clear that the separating hyperplanes do not differ much in this example.

VII. IDENTIFICATION RESULTS

To identify a PWA model of an electronic throttle using the modified clustering-based procedure, \(N = 2600\) data points were collected with sampling time \(T = 5\) ms. The regression vector used is defined by (7), so the model order is \(n = 4\). The parameter \(c\) was fixed on 58 and the number of submodels \(s\) on 12, after a series of trials, having
in mind the tendency towards an outlier free clustering as well as the smallest number of submodels that yields a satisfying model. Scaling factors discussed in Section V are fixed on \( K_\omega = 28 \), \( K_\alpha = 9 \) and \( K_u = 4.5 \). The factors differ from those that normalize the span of each regression vector component by a factor of 1.72, 0.2 and 0.48, respectively. The resulting electronic throttle model consists of 16 affine submodels \( 4 \) of which have been introduced manually to limit the throttle angle respecting its physical constraints. The identification of angle constraints has been avoided on purpose because they are easily retrieved and their implementation in the model is straightforward. On the other hand, identifying other two major process nonlinearities (friction, return spring stress-strain characteristics) is of a much greater importance for controller synthesis. The coefficients of the ARX submodels are printed out in Table I. Coefficients of the polyhedral regions are numerous and aren’t printed out here. The percentage of the regression vectors that are found in their belonging polyhedral region after the regions are calculated is above 92% for all regions. A 3D cut through the identified partition is shown in Figure 4. Figure 5 shows the distribution of regression vectors in the regressor space \( \tilde{X} \) where different symbols mark different submodels. The figure depicts three out of four components of the regression vectors, \( \Theta_{k-1} \), \( \omega_{k-1} \) and \( u_{k-1} \). Submodels marked 1 through 7 are characterized by angle \( \Theta > \Theta_{\text{LH}} \). At the same time submodels marked 1 and 2 are valid for big positive values of angular velocity, submodels marked 3, 4 and 5 are valid for its small absolute values, and those marked 6 and 7 are valid for big negative values of angular velocity (friction remedy). Obviously, the identification procedure successfully recognized different friction regimes. Submodels marked 10, 11 and 12 are characterized by angle \( \Theta < \Theta_{\text{LH}} \) and differ among each other by values of angular velocity. Submodels marked 8 and 9 are valid for angle \( \Theta \approx \Theta_{\text{LH}} \) and angular velocity \( \omega \approx 0 \) for a relatively wide span of input voltage. The second major nonlinearity, the return spring characteristics, is thus also recognized by the procedure. Improvised planes on Figure 5 show the identified hysteretic characteristic of friction and improvised thick lines the identified characteristic of the LH nonlinearity. In addition to the identification, saturation of the throttle angle according to its physical constraints is added manually after the identification procedure ended, by splitting in two the submodels.
that are most likely to overstep the physical limit of the valve plate angle (1, 3, 11, 12), using one-step ahead prediction as a dividing hyperplane in each ($\Theta_k = \text{limit}$).

The validation of the model is performed on the validation dataset by calculating the prediction error $e$ at each step and is depicted in Figure 6. Figure 6 shows that the model is satisfactory in a wide span of throttle angle, with the exception of a very narrow area around $\Theta = \Theta_{\text{LH}}$. This is expected since the switchings of a discrete-time PWA model can only happen in the sampling instants, and thus valid identification in this area would demand more experimental data, more submodels and faster sampling, and the last two demands would be a huge strain for later both pre-computation and implementation of the optimal controller.

VIII. Conclusion

In this work a discrete-time PieceWise Affine (PWA) model of an electronic throttle, a highly nonlinear vehicle’s component, is identified using the clustering-based procedure. As recently shown, for this class of models one can explicitly solve constrained optimal control problems (i.e. compute the control law off-line). The original identification procedure is modified by introducing a linear transformation on the regression vectors prior to the local regression, that enables numerically efficient identification of process nonlinearities. The second modification is introduced in the linear classification algorithm that identifies separating hyperplanes between the clustered data in the regressor space, such that only vertices of each cluster enter the original Multicategory Robust Linear Programming (MRLP). In this way the size of the involved linear program used in MRLP is drastically reduced.

References


Acknowledgments

Thanks to Mato Baotić for his helpful suggestions and discussions and his critical view on the paper and to Giancarlo Ferrari-Trecate and Tobias Geyer who let us use their code for PWA model identification based on clustering.